

IN THE CLAIMS:

Please amend the claims as follows:

- 1 48 (Amended) A speech recognition system comprising:
2 a. an interface coupled to receive a remote session from a speaker; and
3 b. a processing system coupled to the interface to obtain an identification of the
4 speaker and to recognize the speaker's speech wherein the processing system is
5 cumulatively modified according to speech samples obtained during a plurality of
6 remote sessions with the speaker, thereby forming a speaker-specific modified
7 processing system associated with the identification of the speaker.
- 1 50 (Amended) The speech recognition system according to claim 48 wherein the
2 processing system is modified by modifying an acoustic model, thereby forming a speaker-
3 specific acoustic model.
- 1 51 (Amended) The speech recognition system according to claim 50 wherein the
2 processing system includes a memory for storing the speaker-specific acoustic model in
3 association with [an] the identification of the telephone caller.
- 1 52 (Amended) The speech recognition system according to claim 51 wherein the memory
2 stores a plurality of speaker-specific acoustic models, one for each of a plurality of telephone
3 callers and wherein each speaker-specific acoustic model is stored in association with [an] the
4 identification of the corresponding telephone caller.
- 1 53 (Amended) The speech recognition system according to claim 52 wherein the selected
2 ones of the plurality of speaker-specific acoustic models are deleted when a predetermined period
3 of time has elapsed since the corresponding speaker last engaged in a remote session with the
4 voice recognizer.
- 1 54 (Amended) A method of adapting an acoustic model utilized for speech recognition,

wherein the method comprises steps of:

a. obtaining an identification of a speaker;

[a.] b. obtaining a speech utterance from [a] the speaker during a remote session;

[b.] c. recognizing the speaker's speech utilizing an acoustic model during the remote session;

[c.] d. making a determination relative to the speech utterance; and

[d.] e. only when indicated by the determination, performing steps of:

i. modifying the acoustic model according to the speech utterance thereby forming a speaker-specific modified acoustic model; and

ii. storing a representation of the speaker-specific modified acoustic model in association with [an] the identification of the speaker.

57 (Amended) The method according to claim 54 wherein the step of making a determination determines a level of resources available for storing the representation of the speaker-specific modified acoustic model.

REMARKS

Applicant respectfully requests further examination and reconsideration in view of the amendments made above and the arguments set forth fully below. Claims 1-10, 12-26, 28-40, and 42-58 were previously pending in this application. Claims 1-10, 12-26, 28-40, and 42-58 stand rejected. By the above amendments, Claims 48, 50-54 and 57 are amended. Accordingly, Claims 1-10, 12-26, 28-40, and 42-58 are now pending in this application.

Rejections Under 35 U.S.C. § 103

Within the Office Action, Claims 48-58 stand rejected under 35 U.S.C. §103(a) as being unpatentable over U.S. Patent No. 5,127,055 issued to Larkey. The Applicant respectfully traverses this rejection.

Larkey teaches a mechanism for speech recognition within a general use application. The mechanism of Larkey allows for reference patterns to be qualified and stored, where reference patterns are system representations of user speech utterances to be recognized by the system, each reference pattern is assigned a quality which represents how effective that pattern is in enabling the system to recognize an incoming user speech utterance (col. 2, lines 58-2). Specifically, Larkey teaches the digitizing, processing, and analyzing of incoming speech and comparing the incoming speech to reference patterns stored in a reference pattern storage memory (col. 4, lines 10-16). A data processing system then makes a best estimate of the identity of the incoming signal and provides electrical signals identifying the best estimate to an output device (col. 4, lines 16-20). To clarify, "identify" does not refer to the identification of the user; instead, "identify" refers to the recognizing of the incoming user utterance. Larkey further teaches that there may be one or several patterns which correspond to the same word or phrase (col. 4, lines 28-30). The stored reference patterns are dynamically updated and adapted by using correction actions the user has provided about the correctness of the recognition, the correction actions being critical to successful operation of the reference pattern adaptation method (col. 4, lines 32-50). Summarily, Larkey teaches a speech recognition system, to be used as a non-user-specific application, which provides best estimates as to the recognized identity of incoming user utterances, these best estimates are adapted in response to user feedback. *Larkey does not teach*

a user-specific speech recognition system where a speaker-specific model is stored in association with the identification of the speaker and updated according to the incoming user utterance.

In contrast to the teachings of Larkey, the speech recognition system of the present invention provides for speaker-specific acoustic models to be used in the speech recognition process. Multiple users can access the same application, each user having an individualized speaker-specific acoustic model which is stored, retrieved from storage, and modified according to samples of the specific speaker's speech. By utilizing speaker-specific models which are uniquely tailored to the individual user, the speech recognition system of the present invention greatly improves the accuracy of speech recognition over that of a generalized speech recognition system. Further, the speaker-specific acoustic model does not require direct user feedback to be modified. The present invention eliminates the inconvenience of requiring user feedback and as a result improves efficiency by automatically modifying the speaker-specific acoustic models based on the received samples of the specific speaker's speech. Larkey does not teach the use of speaker-specific acoustic models that are modified by samples of the specific speaker's speech.

The independent Claim 48 is directed to a speech recognition system. The system of Claim 48 includes an interface coupled to receive a remote session from a speaker, and a processing system coupled to the interface to obtain an identification of the speaker and to recognize the speaker's speech wherein the processing system is cumulatively modified according to speech samples obtained during a plurality of remote sessions with the speaker, thereby forming a speaker-specific modified processing system associated with the identification of the speaker. The Office Action states that the limitation "the processing system is cumulatively modified according to speech samples" of Claim 48 reads on Larkey "dynamically updated and adapted according to the incoming speech" (D1, col. 4, lines 31-47). The Applicant respectfully traverses this rejection. As discussed above, Larkey teaches that user feedback, in the form of correction actions, must be provided in order for the speech recognition system to be modified. Claim 48 has no such limitation. Specifically, Claim 48 requires that the speech recognition system is modified *according to the sample*. Further, Larkey teaches a speech recognition system, to be used as a *non-user-specific application*, which provides best estimates as to the recognized identity of incoming user utterances, these best estimates are adapted in response to user feedback. *Larkey does not teach a user-specific speech recognition system where a speaker-specific model is stored in association with the identification of the speaker and*

updated according to the incoming user utterance. For at least these reasons, Claim 48 is allowable over the teachings of Larkey.

Claims 49-53 are each dependent upon the independent Claim 48. As discussed above, the independent Claim 48 is allowable over the teachings of Larkey. Accordingly, Claim 49-53 are each also allowable as being dependent upon an allowable base claim.

The independent Claim 54 is directed to a method of adapting an acoustic model utilized for speech recognition. The method of Claim 54 includes the steps of obtaining an identification of a speaker, obtaining a speech utterance from the speaker during a remote session, recognizing the speaker's speech utilizing an acoustic model during the remote session, making a determination relative to the speech utterance, and only when indicated by the determination modifying the acoustic model according to the speech utterance thereby forming a speaker-specific modified acoustic model and storing a representation of the speaker-specific modified acoustic model in association with the identification of the speaker. As discussed above, Larkey teaches that user feedback, in the form of correction actions, must be provided in order for the acoustic model to be modified. Claim 54 has no such limitation. Specifically, Claim 54 requires that the acoustic model is modified *according to the speech utterance*. Further, Larkey teaches a speech recognition system, to be used as a *non-user-specific application*, which provides best estimates as to the recognized identity of incoming user utterances, these best estimates are adapted in response to user feedback. *Larkey does not teach a speaker-specific acoustic model where a speaker-specific acoustic model is modified and stored according to the speech utterance.*

Additionally, as is recognized within the Office Action, Larkey does not teach "storing a representation of the modified acoustic model in association with an identification of the speaker." However, the Office Action states that it would have been obvious to one of ordinary skill in the art to train the system of Larkey, which provides reference patterns which better characterize the speaker's manner of pronouncing a selected word vocabulary, so as to recognize speech and user at the same time. The Applicant respectfully disagrees with this conclusion. It is recognized within the art that speech recognition and speaker identification are separate and distinct technologies. A speech recognition system does not perform speaker identification and a speaker identification system does not perform speech recognition. Therefore it is not obvious that the speech recognition system of Larkey can be trained to perform speaker identification.

The Applicant further contends that even if Larkey is modified as proposed by the Examiner, the result would necessarily constitute a method different from that claimed by the Applicant. Within Claim 54, the speech recognition system is not merely identifying the user but is using that identification to modify and store a speaker-specific acoustic model that correlates directly with the identified user. If by some means Larkey can be adapted to identify the user, as proposed in the Official Action, there is no hint, teaching, or suggestion as to how this identification is to be used, let alone that the identification should specifically be used to modify and store a speaker-specific acoustic model. Though there are no teachings in Larkey to support this position, even if the identification were intended to be used by Larkey in this manner, there is no indication that the speech recognition system as taught by Larkey is capable of performing the speaker-specific modeling and adapting as taught by the present invention. For at least these reasons, Claim 54 is allowable over the teachings of Larkey.

Claims 55-58 are each dependent upon the independent Claim 54. As discussed above, the independent Claim 54 is allowable over the teachings of Larkey. Accordingly, Claim 55-58 are each also allowable as being dependent upon an allowable base claim.

Within the Office Action, Claims 1-10, 12-26, 28-40, and 42-47 stand rejected under 35 U.S.C. §103(a) as being unpatentable over Larkey and further in view of U.S. Patent No. 5,897,616 issued to Kanevsky et al. (hereinafter "Kanevsky"). The Applicant respectfully traverses this rejection.

Kanevsky teaches a mechanism for providing secure access to services and/or facilities using a biometric identification process. The mechanism of Kanevsky allows random questioning, automatic speech recognition, and text-independent speaker recognition techniques to be utilized to *verify the identity of a user*. Specifically, Kanevsky teaches a security system that, through an iterative process, compares user responses to a user database, referred to as speaker candidates, of non-acoustic information and/or an acoustic user model to perform the verification/identification of the user requesting access to a service/facility (Kanevsky, col. 5, lines 34-46). The system first performs an automatic enrollment process by obtaining name, address and whatever other identification is required, this information is referred to as indicia, for building the user model and database used for future identification and verification of the user (Kanevsky, col. 8, lines 23-22). The identification process includes receiving a spoken utterance

containing speaker indicia, decoding the spoken utterance, accessing a database corresponding to a determined speaker candidate of the decoded utterance indicia based on similar indicia within the database, querying the speaker for an additional spoken utterance to obtain additional indicia, receiving and decoding the additional utterance, verifying accuracy of additional decoded utterance indicia against the accessed database, taking a voice sample from the utterances of the speaker and processing the voice sample against an acoustic voice model attributable to the speaker candidate, generating a score corresponding to the accuracy of the decoded answers and closeness of the match between the voice sample and the model, and comparing the score to a predetermined threshold value (Kanevsky, col. 3, lines 22-45). A voice classification module and a text-independent speaker recognition module matches the voice sample of the speaker to voice prints of specific words, the indicia, stored within the accessed acoustic model of the speaker candidate (Kanevsky, col. 6, line 67 to col. 7, line 5; col. 9, lines 46-58; col. 10, lines 56-60). Specific words within the voice sample are determined by a speech recognition device (Kanevsky, col. 6, lines 4-11). Summarily, Kanevsky teaches a system that improves a speaker identification process. More particularly, Kanevsky teaches a system that determines indicia words from a prompted voice sample using a speech recognition device and compares these indicia words to voice prints of corresponding indicia words attributed to a previously identified speaker stored within an acoustic model. *The comparison is made in order to verify the identity of a speaker of the voice sample. Kanevsky does not teach a system to modify a speech recognition system* for the purpose of improving the recognition of indicia words within the voice sample.

Within the Office Action, it is stated that Kanevsky does teach modifying a speech recognition system in association with an identification of the speaker and modifying the speaker-specific speech model. To support this assertion, column 8, lines 16-35 of Kanevsky is cited. The Applicant respectfully disagrees with this reading of those portions of Kanevsky. In lines 16-35 of column 8, Kanevsky teaches that the system provides the ability to automatically adapt, improve or modify its *authentication processes*, and that the authentication process can also include biometrics such as speech and voice prints. As discussed above, the authentication process verifies the identity of an individual based on stored voice prints of specific indicia words. Such verification is not the same as recognizing words within a voice sample. Kanevsky teaches one module for recognizing indicia from within a voice sample (Kanevsky, col. 6, lines

7-11) and another separate module for comparing the recognized indicia words to stored voice prints within an acoustic model for verifying the identity of the speaker of the voice sample (Kanevsky, col. 10, lines 56-60). In fact, Kanevsky teaches using existing speech recognition technology to perform the necessary speech recognition of the indicia words (Kanevsky, col. 13, lines 48-61). It is this type of conventional speech recognition technology that the present invention improves upon (Specification, page 2 lines 3-4; page 2, lines 8-16; and page 2 lines 30-31).

The independent Claim 1 is directed to a method of adapting a speech recognition system. The method of Claim 1 includes the steps of obtaining an identification of a speaker, obtaining a sample of a speaker's speech during a first remote session, recognizing the speaker's speech utilizing the speech recognition system during the first remote session, modifying the speech recognition system according to the sample thereby forming a speaker-specific modified speech recognition system, storing a representation of the speaker-specific modified speech recognition system in association with the identification of the speaker, and using the representation of the speaker-specific modified speech recognition system to recognize speech during a subsequent remote session with the speaker. As discussed above, Larkey teaches that user feedback, in the form of correction actions, must be provided in order for the speech recognition system to be modified. Claim 1 has no such limitation. Specifically, Claim 1 teaches that the speech recognition system is modified *according to the sample*. Further, Larkey teaches a speech recognition system, to be used as a *non-speaker-specific application*, which provides best estimates as to the recognized identity of incoming user utterances, these best estimates are adapted in response to user feedback. *Larkey does not teach a speaker-specific speech recognition system where a speaker-specific model is stored in association with the identification of the speaker and updated according to the incoming speaker sample.* As recognized in the Office Action, Larkey does not teach a “modified speech recognition system in association with an identification of the speaker and modifying the speaker-specific speech model.” However, as discussed above, neither does Kanevsky teach a method of modifying a speaker-specific speech recognition system. Kanevsky teaches a method that improves a speaker identification process. For at least these reasons, Claim 1 is allowable over the teachings of Larkey, Kanevsky, and their combination.

Claims 2-4, 7-10, and 12-16 are each dependent upon the independent Claim 1. As

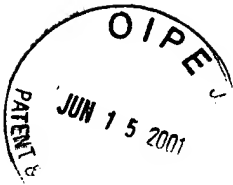
discussed above, the independent Claim 1 is allowable over the teachings of Larkey, Kanevsky, and their combination. Accordingly, Claims 2-4, 7-10, and 12-16 are each also allowable as being dependent upon an allowable base claim.

Further, the Office Action states that the limitation "wherein the representation of the modified acoustic model is a set of statistics which can be utilized to modify a pre-existing acoustic model" of Claim 5 and the limitation "wherein the representation of the modified acoustic model is a set of statistics which can be utilized to modify incoming acoustic speech" of Claim 6 both read on Larkey (col. 2, lines 1-15). The Applicant respectfully traverses this rejection. Larkey teaches that initial reference pattern statistics are established during training of the speech recognition system. However, after initial training, the system adds, deletes, and adapts reference patterns, not statistics (col. 2, lines 16-31). Each reference pattern has associated therewith a quality value representing the effectiveness of that pattern for recognizing an incoming speech utterance (col. 1, line 68 to col. 2, line 3), and each reference pattern stored in memory represents either all or a portion of a word or phrase (col. 2, lines 26-28). Therefore, once the speech recognition system of Larkey is modified, the reference patterns and their associated quality values are used to represent the modified speech recognition system. Larkey does not teach that the modified speech recognition system is represented by a set of statistics. For at least these reasons, Claim 5 and Claim 6 are allowable over the teachings of Larkey.

As an additional reason for allowance, Claims 5 and 6 are each dependent upon the independent Claim 1. As discussed above, independent Claim 1 is allowable over the teachings of Larkey, Kanevsky, and their combination. Accordingly, Claim 5 and 6 are each also allowable as being dependent upon an allowable base claim.

Within the Office Action, Claims 17-26, 28-40, and 42-47 have been rejected as having similar limitations as Claims 1-10 and 12-16. The Applicant respectfully traverses this rejection for at least the same reasons as discussed above pertaining to Claims 1-10 and 12-16.

For the reasons given above, Applicant respectfully submits that all of the remaining claims are in a condition for allowance, and allowance at an early date would be appreciated. Should the Examiner have any questions or comments, he is encouraged to call the undersigned at (650) 833-0160 to discuss the same so that any outstanding issues can be expeditiously resolved.



PATENT
Attorney Docket No.: NUAN-00800

Respectfully submitted,
HAVERSTOCK & OWENS LLP

Dated: 6/14/01

By: Thomas B. Haverstock
Thomas B. Haverstock
Reg. No. 32,571
Attorneys for Applicants

CERTIFICATE OF MAILING (37 CFR § 1.8(a))

I hereby certify that this paper (along with any referred to as being attached or enclosed) is being deposited with the U.S. Postal Service on the date shown below with sufficient postage as first class mail in an envelope addressed to the: Assistant Commissioner for Patents, Washington D.C. 20231

HAVERSTOCK & OWENS LLP.
Date: 6/11/01 By: Carter O'Brien